JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1993, 59, 193-214 NUMBER 1 (JANUARY)

VERBAL SELF-REPORTS ABOUT MATCHING TO SAMPLE: EFFECTS OF THE NUMBER OF ELEMENTS IN A COMPOUND SAMPLE STIMULUS

THOMAS S. CRITCHFIELD AND MICHAEL PERONE

AUBURN UNIVERSITY AND WEST VIRGINIA UNIVERSITY

Adults' self-reports about their choices in a delayed matching-to-sample task were studied as a function of the number of elements (one, two, or three) in a compound sample stimulus. Signal-detection analyses were used to examine control of self-reports by the number of sample elements, by the speed and accuracy of choices reported about, and by several events contingent on self-reports. On each matching-to-sample trial, a sample element appeared as one of two comparison stimuli. Choice of the matching element, if made within 500 ms of the onset of the comparison stimuli, produced points worth money or chances in a drawing for money, depending on the subject. After each choice, subjects pressed either a "yes" or "no" button to answer a computer-generated query about whether the choice met the point contingency. The number of sample elements in the matching-to-sample task varied across trials, and events contingent on self-reports varied across experimental conditions. In Experiment 1, the conditions were defined by different combinations of feedback messages and point consequences contingent on self-reports, but self-reports were systematically influenced only by the sample-stimulus manipulation. Self-report errors increased with the number of sample elements. False alarms (inaccurate reports of success) were far more common than misses (inaccurate reports of failure), and false alarms were especially likely after choices that were correct but too slow to meet the point contingency. Sensitivity (A′) of self-reports decreased as the number of sample elements increased. In addition, self-reports were more sensitive to choice accuracy than to choice speed. All subjects showed a pronounced bias (B'_H) for reporting successful responses, although the bias was reduced as the number of sample elements increased and successful choices became less frequent. Experiment 2 demonstrated that the failure of point contingencies to influence self-reports in the first experiment was not due to a general ineffectiveness of the point consequences. Rates of inaccurate self-reports decreased when they resulted in point losses and increased when they resulted in point gains.

Key words: self-reports, matching to sample, signal detection, sensitivity, bias, button press, button release, adults, college students

Perhaps no single source of data is more widely utilized in psychology than the verbal self-report, which forms the basis of surveys, clinical interviews, many standardized psychological assessments, postexperimental interviews, and several procedures developed to facilitate inferences about private events (e.g., protocol analysis; Ericsson & Simon, 1984). As a matter of practicality, self-reports are a preferred source of data when direct observation or other objective sources of data are unavailable, although few psychologists, regardless of theoretical orientation or area of expertise, appear to view self-reports as universally veridical (e.g., Bem, 1967; Carver & Scheier, 1981; Ericsson & Simon, 1984; Kagan, 1988; Natsoulis, 1988; Nisbett & Wilson, 1977; Shimoff, 1986). Put another way, most psychologists who deal with self-reports acknowledge the possibility of control by variables other than a self-report's putative referents. Such a view is consistent with the behaviorist conception of self-reports as a type of behavior subject to environmental influences (although, as many different theoretical perspectives assert, the "environment" is not limited to the world outside the skin; e.g., see Skinner, 1945, 1957). It follows that some understanding of the variables controlling self-reports should greatly facilitate their interpretation as data.

This was research presented at the 1990 convention of the Association for Behavior Analysis in Atlanta, Georgia. Some data were collected at the Johns Hopkins University School of Medicine while the first author was a postdoctoral fellow under Grant 03389 from the National Institute on Drug Abuse to Roland R. Griffiths. Larry Alferink facilitated collection of the remainder at Illinois State University. We thank Jo Ann Harger for helping run Experiment 2; John Texter, Marcy Pearsall, Scott Gustafson, and Maria Marshall for helping analyze the data; Karla J. Doepke for valuable feedback on drafts of the manuscript; and James A. Dinsmoor for a comment that helped suggest the signal-detection analyses in Figures 3 and 8.
Address correspondence to Tom Critchfield at the Department of Psychology, Auburn University, Auburn, Alabama 36830, or to Michael Perone at the Department of Psychology, West Virginia University, Morgantown, West Virginia 26506-6040.

In principle, then, much could be learned about self-reports as behavior when studied systematically under conditions in which their referents can be corroborated. Ideally, the results of studies following this basic prescription would guide the interpretation of self-reports when corroboration is less practical. Although broad principles probably are not soon to be described, in recent years several behaviorally influenced approaches relevant to this general goal have emerged, with researchers examining self-reports of drug sensations (Overton, 1987), reinforcer value (Bernstein & Michael, 1990), reinforcement contingencies (Catania, Matthews, & Shimoff, 1982; Wasserman & Neunaber, 1986), alcohol consumption (Tucker, Vuchinich, Harris, Gavornik, & Rudd, 1991), and children's play activities (de Freitas Ribeiro, 1989). In addition, a number of studies have focused on responses by animals that appear to share characteristics with humans' self-reports (e.g., Shimp, 1983; Ziriax & Silberberg, 1978).

We have described an approach designed, for laboratory convenience, on two precepts (Critchfield & Perone, 1990a, 1990b). First, in developing the procedure we sought to keep the putative referent—the target behavior—under experimental control so its characteristics might be subject to manipulation. The target behavior involved a delayed matching-to-sample (DMTS) task in which reinforcement depended on selecting the correct stimulus within a limited time. Success, measured as the percentage of responses meeting the reinforcement contingency, is known to vary systematically as a function of both the time limit and attributes of the sample and comparison stimuli (e.g., see Baron & Menich, 1985a, 1985b), so the procedure generates target behavior that can be manipulated in different ways. Second, we sought to restrict the number of possible self-reports to aid in interpreting the degree of correspondence between the reports and the target behavior. Subjects reported the success of each DMTS response by pressing "yes" or "no" buttons in answer to a computer-generated query (e.g., asking whether that last target response met a reinforcement contingency).

In our previous studies, the target behavior was varied by manipulating the stringency of the time limit. Self-report accuracy generally was lower when time limits on target responding were more stringent. The purpose of the present investigation was to address an issue left unresolved by that finding. When a time limit on the target response influenced self-report accuracy, it was not clear whether the critical variable was the time limit per se or its effect on target-response success. That is, when time limits were stringent, target responses predictably met the reinforcement criteria less often. Thus, the time pressure and its effect on target behavior were confounded.

In the present investigation, target-response success was manipulated in a different way. On different trials, subjects viewed a sample stimulus with one, two, or three elements. Only one of these elements appeared among the two comparison stimuli. The time limit on selecting a comparison stimulus remained constant. Using similar contingencies and manipulations in a study not including self-reports, Baron and Menich (1985b) found that the percentage of successful DMTS response decreased as the number of sample elements increased. Thus, this within-session manipulation allowed us to determine whether self-report accuracy was related to DMTS success rates when manipulated via stimulus characteristics rather than the time limit on target responding.

The experimental conditions were defined by different types of consequences and feedback that were contingent on the self-reports. Because previous laboratory studies have shown contingent reinforcement to influence the self-reports of children (de Freitas Ribeiro, 1989; Risley & Hart, 1968), it seemed reasonable to assume that similar contingencies would influence the self-reports of adults. Five experimental conditions involved the presence or absence of (a) feedback about the point-reinforcement outcome for the preceding DMTS target response, (b) feedback about the correspondence of the self-report to the target response, and (c) a point bonus or penalty contingent on accurate or inaccurate self-reports, respectively.

EXPERIMENT 1

METHOD

Subjects

Six men, ranging in age from 18 to 46 years, volunteered to participate in a laboratory ex-

Table 1
Summary of subject characteristics and compensation plans.

Subject	Age	Vocation	Base wage	Point value
1	29	Unemployed	$3.00 per hour	0.5 cents
2	46	Postal employee	$3.00 per hour	0.5 cents
3	29	Unemployed	$3.00 per hour	0.5 cents
4	29	Undergraduate student	Bonus credit in classes	1 lottery chance
5	18	Undergraduate student	Bonus credit in classes	1 lottery chance
6	39	Unemployed	None	1 cent

periment on "Human Performance and Decision-Making." Subjects 1, 2, and 3 were community members recruited through newspaper advertisements. Subjects 4 and 5 were undergraduate students recruited through a psychology department subject pool. Subject 6 was a recreational sedative abuser who was living in a residential research unit as part of another long-term investigation involving drug effects. He participated in the present study to earn extra money on days when his presence was required at the research facility but the drug study was not operating. Aside from Subject 6's drug use, the men all reported that they were free of current medical or psychological problems. They all signed an informed consent agreement that explained compensation procedures. Subjects 1, 2, and 3 earned a $3.00 hourly wage supplemented by earnings during experimental sessions, for an average wage of about $8.00 per hour. Subjects 4 and 5 earned bonus credit in psychology classes based on their hours of participation; during sessions they accumulated points that served as chances in a lottery for cash prizes. Subject 6 earned no base wage (he was paid a daily wage by the residential research unit) but earned about $8.00 per hour during experimental sessions. Table 1 summarizes subject characteristics and compensation procedures.

Apparatus

Each man worked alone in a small booth (3 m by 3 m) containing a table, chair, and a response console with a monochrome video monitor resting on it (for details, see Critchfield & Perone, 1990b). The men performed the DMTS task using red push buttons (with a surface area about 1 cm square) mounted to a small box extending from each side of the console. Two white lamps near the top of the console's sloping front panel signaled point reinforcers in some conditions. Self-reports were made using two 3-cm round back-illuminable response keys arranged horizontally about 12 cm from the bottom of the console's front panel. During work periods, the men wore sound-muffling earphones. A microcomputer in an adjacent booth controlled experimental events and collected the data.

General Procedure

Trials. During the main experiment each trial consisted of one DMTS response followed immediately, in sequence, by (a) a self-report, (b) feedback about the success of the DMTS response (if scheduled), and (c) feedback or consequences contingent on the self-report (if scheduled). Trials were separated by an intertrial interval (ITI) lasting at least 1 s. As described below, subjects initiated each trial at the end of the ITI. This ensured that a subject was oriented toward the video screen when stimuli were presented, but also meant that the ITI could extend beyond its nominal value.

The video screen was divided into an upper box, used to display stimuli and messages relevant to the DMTS task, and a lower box, used to present messages relevant to the self-report portion of the trial. A trial began with the instruction "HOLD RED BUTTONS DOWN" displayed in the center of the upper box. Depressing both side buttons cleared the message and produced a 1-s sample-stimulus display in the center of the box (these side buttons remained depressed until used to select a comparison stimulus). The stimuli are described below. Following a 1-s delay, two comparison stimuli appeared, one 2 cm to the left and one 2 cm to the right of the center of the box. One comparison stimulus was randomly generated and the other matched a sample element. Release of the red button on the same side as the matching element was counted as correct. A point reinforcer was earned if a correct response occurred less than 500 ms after presentation of the comparison stimuli.

Immediately after the choice response, the DMTS box on the video screen cleared and the center of the self-report box displayed the query, "Did you score?" (As noted below, the word "score" was used to signal point delivery during preliminary training.) Below it were the words "YES" and "NO" printed 1 cm from the right and left sides of the self-report box, respectively. Simultaneously, the two self-report buttons were illuminated. Pressing either button cleared the screen and advanced the trial to the next scheduled event. Other buttons were ineffective. This method of self-reporting, although artificial by traditional linguistic standards, appears to meet definitional criteria for verbal behavior specified in at least one behavioral account (Skinner, 1957).

When scheduled, feedback about the success of the DMTS response immediately followed the self-report. During the experiment proper, a correct DMTS response within the time limit produced the 1-s message, "You scored! *x* point[s] added to your earnings" (x = 1 or 2, depending on the experimental condition; see Table 3 and related text below). Incorrect or late responses produced the message "No score," also for 1 s. When no DMTS feedback was scheduled, the trial advanced immediately to the next event.

When scheduled, feedback about the accuracy of the self-report or point consequences contingent on self-report accuracy followed next. The relevant messages are included in the description of experimental conditions.

Throughout, error messages discouraged responses not conforming to the experimental protocol. For example, release of DMTS buttons before comparison stimuli were presented produced a message stating, "Illegal Action!" and caused the trial to begin again (for further details, see Critchfield & Perone, 1990b).

Sessions. From trial to trial, the number of elements in the sample stimulus varied from one to three according to an irregular sequence that was randomly selected from a set of 36 sequences. Each sequence was arranged with an equal number of trials per session with one, two, and three sample elements; each also limited consecutive repetitions of any type of sample to three trials. Sessions consisted of 96 trials divided into two 48-trial blocks separated by a 20-s intermission, during which the screen was blank except for a message stating, "Intermission—Please wait." Sessions typically lasted 10 to 15 min, and usually eight sessions were conducted in about 2 hr during each visit to the laboratory, allowing for brief subject-initiated rest periods between the sessions. At the end of each session, a message on the subject's screen displayed the number of points accumulated during that session.

Stimuli. Each stimulus element consisted of a 6 × 3 matrix of rectangular cells, of which as few as 3 or as many as 18 could be illuminated (similar stimuli were described by Baron & Menich, 1985b). An element could be as large as 10 mm by 7 mm, depending on how many cells were illuminated. On each trial, stimulus elements were drawn randomly from a pool of several thousand unique shapes, without replacement except for the obvious exception that one sample element matched one comparison stimulus. On one-element trials, a single sample element was displayed in the center of the DMTS box; on two- and three-element trials, multiple sample elements were spaced 1 cm apart, with the entire sample-stimulus array centered inside the DMTS box.

General Instructions

Before the first session, the men read printed instructions covering the following points (quotations indicate exact phrasing; other portions are paraphrased for brevity): (a) "Your job is to make decisions based on information presented on your computer screen, and to indicate your decisions using buttons on the console." (b) "When you depress the red buttons, one or more *sample* shapes will appear on your screen, then disappear. Shortly thereafter, two *test* shapes will appear. You should indicate which of these test shapes matches one of the sample shapes." (c) You can earn a point each time you choose the correct (matching) test shape. In order to earn a point, your choice must be both correct and within a time limit. (d) "Sometimes messages or questions will follow your choice of a test shape. The basic decision-making procedure remains the same regardless of what happens after your selection." (e) Do not attempt to ask questions or leave the room during a session. (f) "Beyond the information contained in these instructions, it is up to you to decide how to operate

Table 2
Summary of experimental manipulations.

Programmed event	Experimental condition A	B	C	D	E
DMTS point contingency	2 points per score	2 points per score	2 points per score	1 point per score	1 point per score
DMTS feedback	yes	no	no	no	no
Self-report feedback	no	no	yes	yes	no
Self-report contingency	no	no	no	acc = +1 point inacc = −1 point	acc = +1 point inacc = −1 point

Note. DMTS = delayed matching to sample, acc = accurate, inacc = inaccurate.

the console to your best advantage. You may do what you like during sessions but remember that your point earnings depend on what you do."

Additional instructions, described below, accompanied the introduction of the experimental conditions.

Preliminary Training

Each man participated in two preliminary training phases, each lasting eight sessions and consisting solely of DMTS trials without self-reports. In the initial training phase, following DMTS response, three feedback messages were simultaneously displayed for 2 s. The first message stated, "Your choice was CORRECT [or WRONG]." The second message stated "Your choice was FAST ENOUGH [or TOO SLOW]". The final message summarized the implications of the other message for point reinforcement, stating either "YOU SCORED! 2 points added to your total," or "NO SCORE." At the beginning of this phase, the time limit for responding was 2,000 ms and decreased across blocks of 48 trials according to the following sequence: 1,250, 1,000, 850, 750, 700, 650, 600, 550, and 500 ms. Thus, by the end of the fifth session the time limit had reached its final value for the experiment (500 ms). The subject's printed instructions did not describe this gradual reduction in the time limit.

The second training phase was identical to its predecessor, with three exceptions. First, the time limit remained stable across sessions at 500 ms, as it did during the remainder of the study. Second, the feedback display was simplified so that subjects no longer received feedback specifically describing their speed and accuracy. A successful DMTS response produced only the message, "YOU SCORED! 2 points added to your total." An unsuccessful response produced the message, "NO SCORE." Third, display time for the feedback message was reduced from 2 s to 1 s.

Experimental Conditions

Following preliminary training, each man participated in five experimental conditions in which a self-report was prompted immediately after each DMTS response. The conditions differed from one another in terms of the presence or absence of (a) the feedback message indicating DMTS success, (b) an additional point contingency on self-report accuracy, and (c) a feedback message indicating self-report accuracy. Table 2 summarizes the five experimental conditions with respect to these features, as described in detail below. In every case, however, points remained contingent on DMTS responses that were correct and faster than the time limit, regardless of whether feedback was provided after each trial. Table 3 shows that the conditions were presented in a different sequence for each man, and that some conditions were replicated for some subjects as time permitted.

Table 3
Sequence of experimental conditions for each subject.

Subject	Position in sequence 1	2	3	4	5	6	7	8	9
1	A	B	C	D	E	D			
2	B	A	B	C	E	D	C	—[a]	D
3	D	B	A	C	E				
4	E	C	B	D	A				
5	C	D	E	A	B				
6	B	A	D	C	E	D			

[a] Condition 8 for Subject 2 did not involve self-reports and is not described.

At the beginning of each condition, subjects read a card describing the point contingencies. In Conditions A through C, the card stated, "Today you can earn points from: Matching—2 points for each score." In Conditions D and E, the card read "Today you can earn points from: Matching—1 point for each score; Self-Reports—1 point bonus for each accurate report, 1 point penalty for each inaccurate report." The card stated that point consequences applied regardless of whether they were signaled on the screen, and remained in the experimental room for reference throughout the condition.

Condition A (DMTS feedback). Reporting was followed by the DMTS feedback message. No feedback specifically described self-report accuracy. Two points were contingent on DMTS responding, and no points were contingent on self-report accuracy.

Condition B (no feedback). Reporting was followed immediately by the intertrial interval. No feedback was provided about either DMTS performance of self-report accuracy. Two points were contingent on DMTS responding, and no points were contingent on self-reporting.

Condition C (self-report feedback). Reporting was followed by a 1-s message in the self-report box describing the accuracy of the self-report, but not specifically the outcome of the DMTS response. The message stated "Results of Your Report" followed by "CORRECT" or "WRONG." Two points were contingent on DMTS responding, and no points were contingent on self-report accuracy.

Condition D (self-report contingency plus feedback). Reporting was followed by a 1-s message in the self-report box describing the accuracy of the self-report and the outcome of a contingency on report accuracy. Accurate reports were followed by, "Results of Your Report," and, "CORRECT—1 point bonus." Following an inaccurate self-report, the latter message stated, "WRONG—1 point penalty." Thus, 1 point was contingent on DMTS responding, and 1 point (plus or minus) was contingent on self-report accuracy.

Condition E (self-report contingency, no feedback). Reporting was followed immediately by the intertrial interval. No feedback messages described either DMTS performance or self-report accuracy. One point was contingent on DMTS responding, and 1 point (plus or minus) was contingent on self-report accuracy.

Each condition lasted eight sessions. The expectation, based on pilot work, that performance would stabilize over this interval generally was borne out. Overall percentages of DMTS success (responses that met the point contingency) and self-report accuracy from the final five sessions per condition were collapsed across trial type and subjected to the following post hoc stability test: The difference between mean percentages in the first and last two sessions was considered as a proportion of the five-session grand mean. For DMTS success, this proportion was less than .10 in 80% of the cases (less than .15 in 91% of the cases). For self-report accuracy, the proportion was less than .10 in 83% of the cases (less than .15 in 94% of the cases).

RESULTS

Results are based on the last five sessions (480 trials) of each condition. The between-conditions manipulations of feedback and point consequences produced no systematic effects, and will be mentioned below only where appropriate. For this reason, and for clarity of data presentation, within-subject replications of experimental conditions were summed prior to data analysis (Table 3). Thus, some functions represent more than 480 trials.

DMTS Performance

Figure 1 shows DMTS success in terms of the percentage of responses that met the conjunctive contingency on speed and accuracy. The figure includes one panel for each feedback-consequence condition. Each panel shows performance of the 6 subjects as a function of the number of DMTS sample stimuli. Typically, then, each data point represents 160 self-reports representing 32 trials per session with each type of sample (one, two, or three sample elements), summed over five sessions.

Increasing the number of sample elements reduced DMTS success. Success rates usually were higher than 85% on one-element trials and in the range of 36% to 75% on three-element trials, with 27 of 30 functions showing monotonically decreasing trends.

Overall Correspondence of Self-Reports to DMTS Performance

Figures 2 and 3 are organized in the same format as Figure 1, with each panel summarizing the performance of 6 subjects in one feedback-consequence experimental condition.

	EXPERIMENTAL CONDITIONS A	B	C	D	E
DMTS Feedback	X				
Self-Report Feedback			X	X	
Self-Report Contingency				X	X

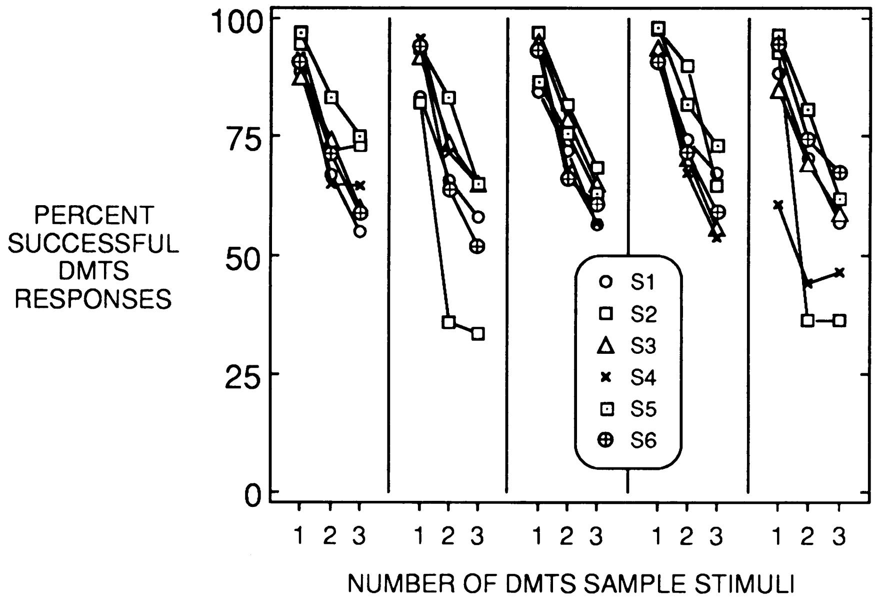


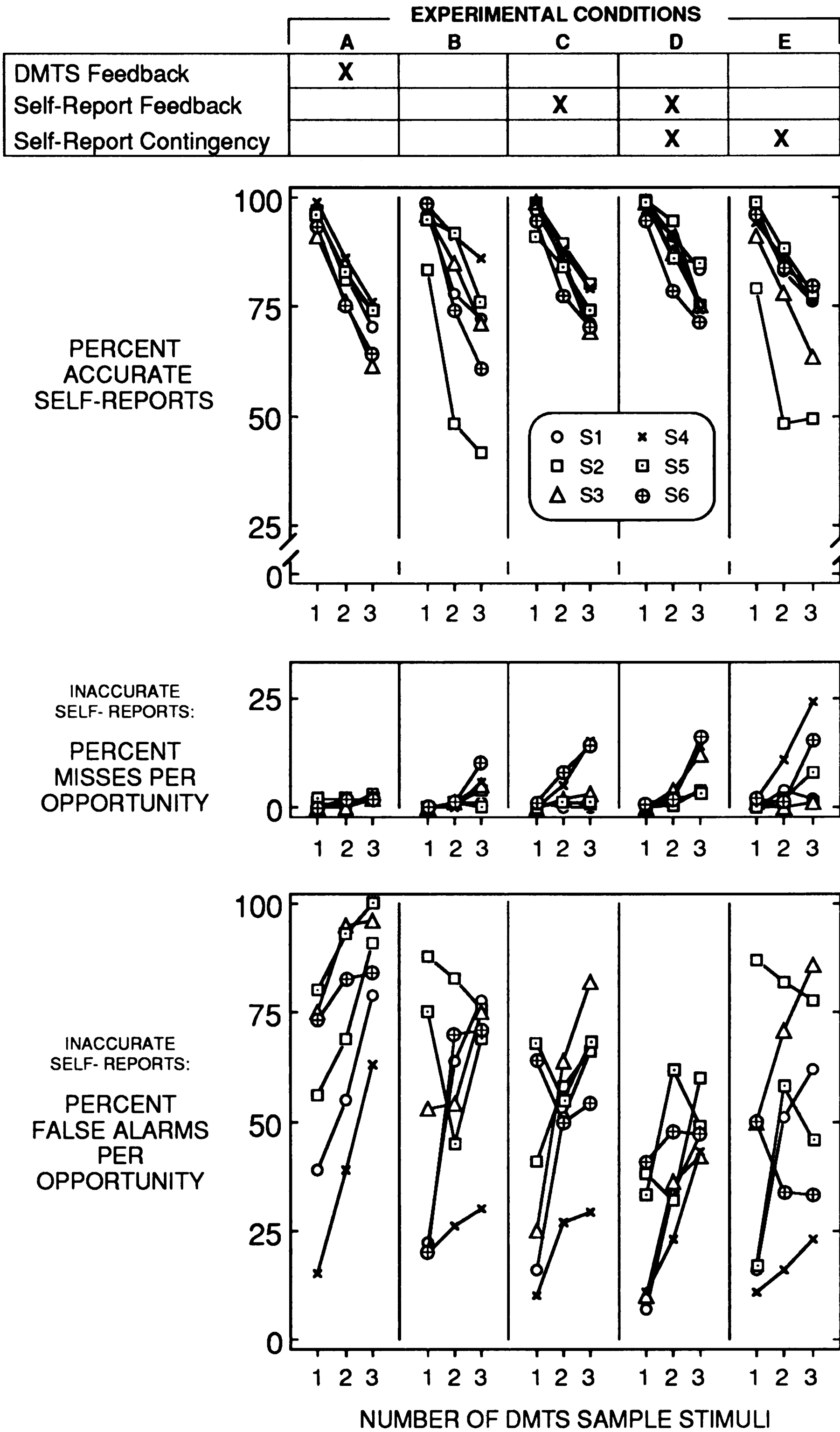
Fig. 1. Experiment 1: Percentage of responses in the DMTS target task that met the conjunctive (accuracy plus speed) reinforcement contingency. Each panel shows performance of 6 subjects as a function of the number of DMTS sample stimuli in the last five sessions of one experimental condition. Note that some functions represent performance collapsed across replications of a condition (see text and Table 3).

The top portion of Figure 2 shows the percentage of accurate self-reports. Reports following one-element trials almost always were accurate, and accuracy tended to decrease as the number of elements increased (29 of 30 functions showed monotonically decreasing trends). The remaining two panels present an error analysis of the self-reports, using terminology borrowed from signal-detection analysis (Green & Swets, 1966). Considering an accurate DMTS response within the time limit as the "signal," an inaccurate self-report could consist of a *miss* (a report of failure following a successful response) or a *false alarm* (a report of success following an unsuccessful response).

Because a miss could occur only after a successful response, and a false alarm could only follow an unsuccessful one, the middle and bottom portions of Figure 2 present rates of misses and false alarms, respectively, per opportunity. Misses were relatively rare, usually occurring on fewer than 15% of opportunities. By contrast, false alarms occurred on a larger percentage of opportunities. Thus, when the men reported inaccurately, they were likely to do so by falsely reporting a successful response.

Miss rates, although low, appeared to increase slightly with the number of DMTS sample elements, although trends may be difficult to detect because of possible floor effects. Although miss rates increased monotonically in only half the functions, in 26 of 30 cases miss rates were higher for three-element trials than for one-element trials.

False alarm rates were variable, but showed a tendency to increase with the number of DMTS sample elements. Considering all five conditions, a monotonically increasing trend appeared in 20 of 30 cases (three-element rates exceeded one-element rate in 24 of 30 cases), with considerable differences across subjects. For example, functions were always monotonic for Subjects 1, 3, and 4, but were mono-


EXPERIMENTAL CONDITIONS
A
B
C
D
E
DMTS Feedback
Self-Report Feedback
Self-Report Contingency
X
PERCENT ACCURATE SELF-REPORTS
S1
S2
S3
S4
S5
S6
INACCURATE SELF-REPORTS:
PERCENT MISSES PER OPPORTUNITY
INACCURATE SELF-REPORTS:
PERCENT FALSE ALARMS PER OPPORTUNITY
100
75
50
25
0
1 2 3
NUMBER OF DMTS SAMPLE STIMULI

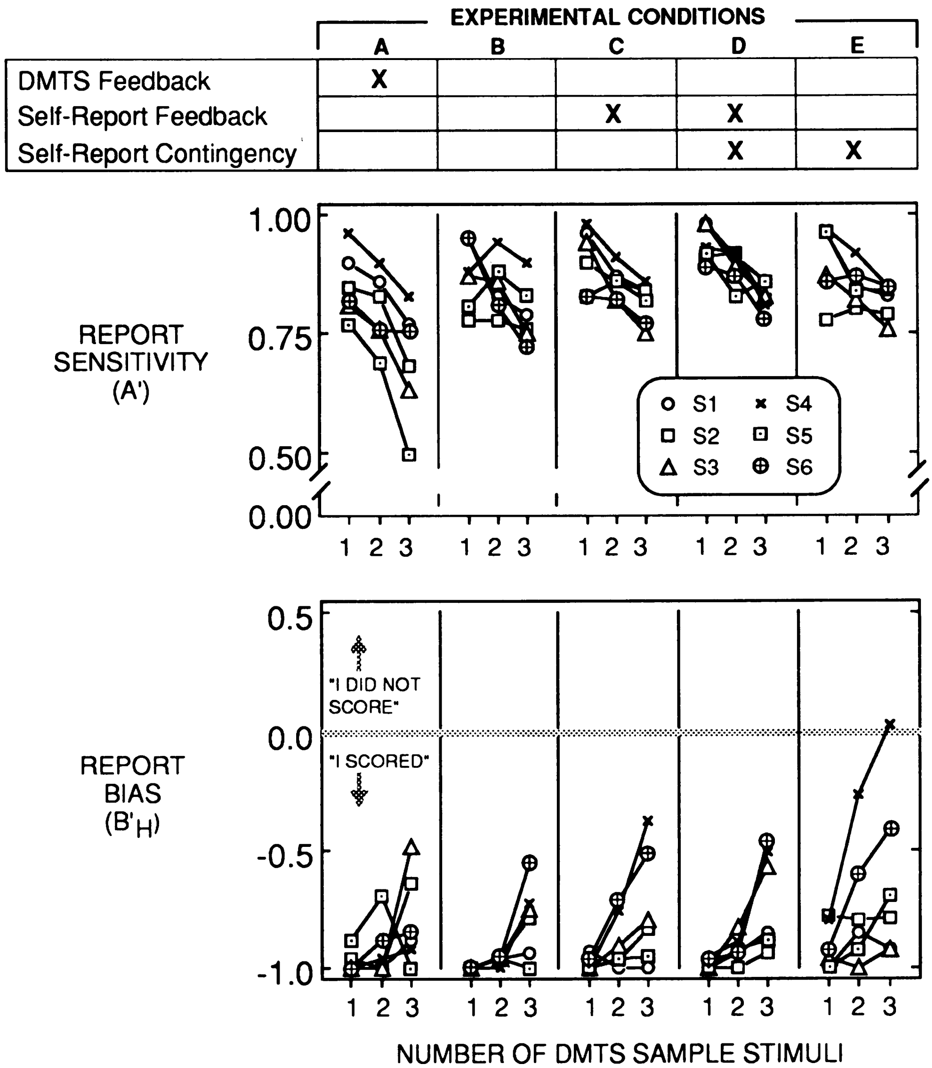


Fig. 3. Experiment 1: Report sensitivity and report bias as a function of the number of sample stimuli in the DMTS target task. Each panel shows performance of 6 subjects in the last five sessions of one experimental condition. Note that some functions represent performance collapsed across replications of a condition (see text and Table 3).

tonic only in one of five conditions for Subject 5. Comparing across conditions, intersubject variability was smallest in Condition D (in which self-reports produced point consequences and feedback messages), due to the absence of the extremely high rates of false alarms seen in other conditions. However, the fact that not all subjects showed decreased rates of false alarms in this condition precludes any assumption of a general effect.

The signal-detection analogy implied by Figure 2 was extended by calculating non-parametric indices of sensitivity and bias (Grier, 1971) using rates of false alarms and

←

Fig. 2. Experiment 1: Self-report accuracy and rates of inaccurate self-reports in two categories (misses and false alarms) suggested by signal-detection analysis. Each panel shows the accuracy of 6 subjects as a function of the number of DMTS sample stimuli in the last five sessions of one experimental condition. Note that some functions represent performance collapsed across replications of a condition (see text and Table 3).

hits (hit = report of success after a successful response; hit rates are the difference between 1 and the miss rate). Figure 3 presents the results of these analyses. The top panel shows report sensitivity scores. For the present data set, the A′ measure of sensitivity (Grier, 1971, Formula 2) estimates an individual's detection of a signal consisting of a successful DMTS response. Values can range from 1.0 to 0, with .50 indicating chance levels of correspondence between self-reports and DMTS outcomes. In most cases, A′ decreased when the number of DMTS element increased (monotonic trend in 23 of 30 cases, and one-element sensitivity exceeding three-element sensitivity in 27 of 30 cases).

The bottom panel of Figure 3 shows report bias scores. For the present data set, the B'_H measure of bias (Grier, 1971, Formulas 7 and 8) estimates an individual's tendency to report successful or unsuccessful DMTS responses, regardless of actual DMTS performance. Values can range from −1.0 to +1.0, with negative values representing a bias toward reporting successful target behavior and positive values representing a bias toward reporting unsuccessful target behavior. The most salient feature of the bias functions is their location near the negative end of the scale, indicating a pervasive bias for reporting successful responses. There was also a tendency for bias to become less extreme as the number of sample elements increased. In this regard, 17 of 30 functions were monotonic, and in 26 of 30 cases the bias was less extreme for three-element trials than for one-element trials.

Discussion

Two previous studies found that self-report accuracy decreased when a time limit on the DMTS target behavior was made more stringent (Critchfield & Perone, 1990a, 1990b). The present study was designed to determine whether a different type of challenge to DMTS performance would have similar effects. The within-session sample-stimulus manipulation controlled target-task success reasonably well, as evidenced by the negative covariation of DMTS success rates with sample-element number. More important, essentially the same relation was observed between self-report accuracy and element number; that is, as sample elements became more numerous, self-reports tended to become less accurate.

Signal-detection terminology and indices permitted a more precise description of self-reporting patterns than was possible using the more global measure of report accuracy (for related applications, see Appel & Dykstra, 1977; Gardner, Martinez, & Espinoza, 1987; Hosseini & Ferrell, 1982; Zuroff, Colussy, & Wielgus, 1983). In particular, examination of inaccurate self-reports showed that, per opportunity, false alarms were more likely than misses. That is, when the men reported DMTS performance inaccurately, they tended to overestimate their success. This pattern was conveniently summarized with the bias index, which permitted quantification of a tendency we have described previously in more general terms (Critchfield & Perone, 1990a, 1990b). This in turn showed that in many cases bias tended to become less extreme as DMTS sample elements became more numerous. Sensitivity scores appeared to be related to the number of sample elements as well.

EXPERIMENT 2

In Experiment 1, an inverse relationship between self-report accuracy and the number of DMTS sample stimuli was obtained for three types of subjects (students, community members, and a drug abuser) under three different compensation systems. However, conspicuously absent were any systematic effects of feedback and point consequences contingent on self-reports. It is possible that more powerful consequences would have produced different effects. Yet the points that subjects earned for accurately self-reporting represented three different consequences for different subjects (lottery chances and two different cash values; see Table 1). If magnitude of the consequences was the sole important factor, one might expect to see consequence-based differences in reporting patterns across subjects, but none were apparent. Nevertheless, it seemed reasonable to conduct a control experiment to determine whether the consequences of self-reporting, as scheduled in Experiment 1, were capable of altering patterns of self-reporting.

The second experiment sought to determine whether the consequences of self-reporting used in Experiment 1 would be effective if scheduled in a different way. One experimental condition replicated the procedures of Condition D from the previous experiment, so that accurate self-reports produced a point gain and inaccurate self-reports produced a point loss. A second inverted the contingencies so that accurate reports produced a point loss and inaccurate reports produced a point gain. If the point contingencies were completely irrelevant to patterns of self-reporting, this manipulation would be expected to have no effect on self-report accuracy.

Method

Subjects and Apparatus

Three undergraduate students, aged 18 to 21, were recruited through a psychology department subject pool. Subjects 7 and 8 were female and Subject 9 was male. They received bonus credit in psychology courses based on the number of hours they participated, and during experimental sessions they accumulated points that counted as chances in a lottery for cash prizes. The apparatus was the same as in Experiment 1.

Procedure

As in the previous experiment, a trial consisted of a DMTS response, a self-report, feedback messages, and an intertrial interval. Sessions consisted of 96 trials, 32 each with one, two, and three sample elements. Session earnings were displayed on the video screen at the end of each session. Conditions generally lasted eight sessions. General instructions and pretraining procedures were identical to those used in Experiment 1, as were experimental procedures, except as specified below. Because of scheduling difficulties, preliminary training for Subjects 8 and 9 lasted eight sessions instead of 16.

Experimental Conditions

Each subject participated in two experimental conditions. As in Condition D of Experiment 1, subjects could earn up to 2 points on each trial. One point was contingent on a correct DMTS response within the time limit. A 2nd point (gain or loss) was contingent on the self-report. The two conditions differed in terms of the specific self-report contingency.

In the "truth" condition, an accurate self-report produced the message "Results of Your Report: 1 POINT BONUS." An inaccurate self-report produced the message, "Results of Your Report: 1 POINT PENALTY." In the "lie" condition, an accurate self-report produced the penalty message, and an inaccurate one produced the bonus message. Before the first experimental session subjects read the following printed instructions, which remained posted in the work booth throughout all experimental conditions:

> Each time you make a correct match within the time limit, you will "score" 1 point. However, you will not receive feedback about the success of each matching response. In addition, after each attempted match, the computer will ask you whether you believe you scored. If your answer is acceptable you will earn a 1 point bonus. If your answer is unacceptable, a 1 point penalty will be subtracted from your total.

The truth condition (A) and lie condition (B) alternated in A-B-A-B sequence for Subjects 7 and 8, and in B-A-B-A sequence for Subject 9.

Perhaps because of incomplete pretraining for 2 of the subjects, performances achieved lower levels of stability than in Experiment 1. According to the post hoc test used in the previous study, the index of stability for percentages of reinforced DMTS trials and accurate self-reports was within .15 in 75% of the cases. The less stable performances occurred almost exclusively for Subjects 8 and 9 during the first two experimental conditions.

Results and Discussion

Analyses are based on the last five sessions of a condition. Figure 4 (top) shows the percentage of DMTS responses that were successful in meeting the conjunctive accuracy-speed point contingency. Each panel shows the DMTS performance of 3 subjects as a function of the number of DMTS sample stimuli in one experimental condition defined by the contingency on self-reports. Within each panel it is apparent that, as in Experiment 1, DMTS success tended to become less frequent as the number of sample stimuli increased. Across panels, no systematic effects of the truth versus lie manipulation were apparent. Subject 7's DMTS performance was similar across conditions, whereas Subjects 8 and 9 showed gradual increases in success rates across conditions, probably due to insufficient preliminary training.

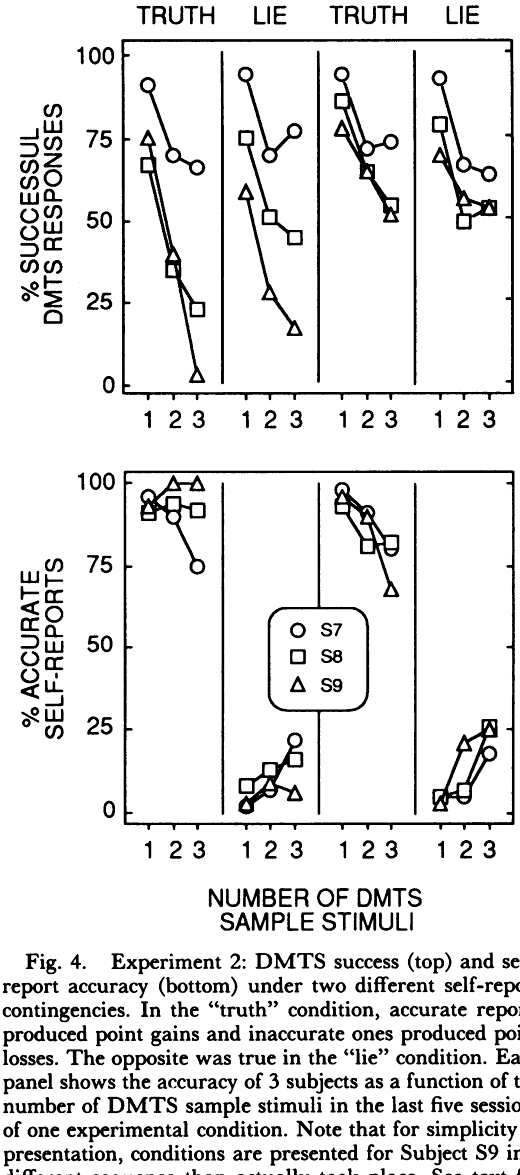


Fig. 4. Experiment 2: DMTS success (top) and self-report accuracy (bottom) under two different self-report contingencies. In the "truth" condition, accurate reports produced point gains and inaccurate ones produced point losses. The opposite was true in the "lie" condition. Each panel shows the accuracy of 3 subjects as a function of the number of DMTS sample stimuli in the last five sessions of one experimental condition. Note that for simplicity of presentation, conditions are presented for Subject S9 in a different sequence than actually took place. See text for details.

The bottom portion of Figure 4 shows the percentage of accurate self-reports in each condition as a function of the number of DMTS sample element. As in previous figures, each panel shows the performance of individual subjects as a function of the number of sample elements in one experimental condition. Accuracy was high in the truth conditions and low in the lie conditions. Figure 5 shows that this general effect was reflected in rates of the two types of reporting errors. Each panel shows miss and false alarm rates, per opportunity, for one trial type across the four experimental conditions. In all cases, these rates were relatively low in the truth condition and high in the lie condition.

In the present study, then, the same combination of point consequences and outcome feedback that appeared to be ineffective in Experiment 1 was sufficient to alter patterns of self-reporting when scheduled in a different way (i.e., the lie condition). These results were obtained using presumably the weaker back-up reinforcer from Experiment 1 (lottery chances vs. money). It cannot be concluded, therefore, that the absence of experimental effects across conditions in the first experiment was due to a globally weak set of manipulations.

This does not, however, answer the question of why the same type of contingency was successful in decreasing (Experiment 2), but not increasing (Experiment 1), self-report accuracy. Certainly under other circumstances feedback and reinforcement contingencies have improved the correspondence between self-reports and their behavioral referents in humans (e.g., Caracci, Mukherjee, Roth, & Decina, 1990; Ciminero, Nelson, & Lipinski, 1977; de Freitas Ribeiro, 1989; Hefferline & Perera, 1963; Risley & Hart, 1968), and such contingencies are necessary even to establish self-reporting responses in animals (e.g., Shimp, 1983). It is possible that ceiling or floor effects limited the influence of feedback and contingency manipulations in Experiment 1, but across three trial types and multiple dependent variables (e.g., accuracy, rates of misses and false alarms) there were many instances in which room for improvement was apparent.

The results of Experiments 1 and 2 might not appear inconsistent to theorists who believe that some aspects of self-knowledge occur "automatically" or not at all. From this perspective, in cases in which self-information is not automatically conscious, self-reports cannot be enhanced by manipulations of the sort used in Experiment 1 (Ericsson & Simon, 1984; Hasher & Zacks, 1979). Yet conscious self-information does not guarantee accurate self-reporting, because the reports themselves can still be distorted by external influences, as in Experiment 2 (e.g., Nisbett & Wilson, 1977).

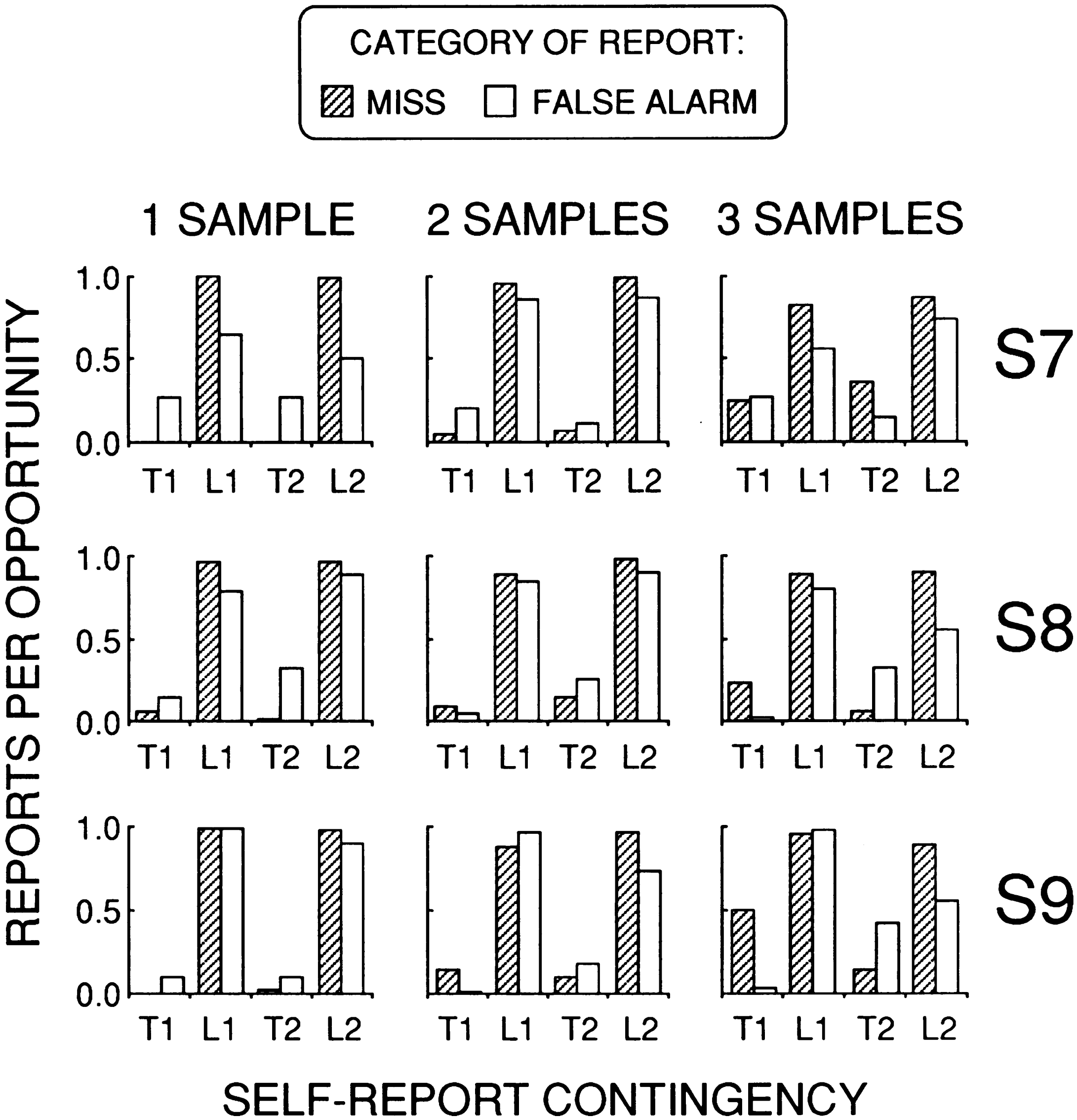


Fig. 5. Experiment 2: Rates of two categories of inaccurate self-reports, misses and false alarms, under two different self-report contingencies. In the "truth" condition (T1 and T2), accurate reports produced point gains and inaccurate ones produced point losses. The opposite was true in the "lie" condition (L1 and L2). Each panel shows rates for one trial type (based on the number of DMTS sample stimuli) for 1 subject across experimental conditions. Note that for simplicity of presentation, conditions are presented for Subject S9 in a different sequence than actually took place. See text for details.

The relevance of automaticity hypotheses to the present research may be limited due to a reliance on variables that cannot be observed or manipulated (Hayes, 1986). Nevertheless, such hypotheses may be valuable in reminding us that different variables can, on occasion, be responsible for accurate and inaccurate self-reports. In the present Experiment 1, self-reports were reasonably accurate regardless of whether point contingencies were in effect. In Experiment 2, therefore, the distortion of self-reports by point contingencies (in the lie condition) must represent something other than the inversion of contingencies normally responsible for accurate reporting. Rather, point consequences may have merely redefined the functions of the two report keys while leaving the stimulus control of target behavior over reporting responses unchanged. Note in Figure 4 the tendency for accuracy functions in the truth and lie conditions to appear as mirror images of one another. These patterns would be expected if the lie contingency served to redefine the "yes" report button as a "no" response and the "no" report button as a "yes" response, without influencing a subject's discrimination of actual DMTS performance.

Finally, it should be noted that the absence of systematic effects across feedback and consequence conditions in Experiment 1 represents a failure to replicate one aspect of our previous findings. Critchfield and Perone (1990b) reported that self-reports of DMTS were more accurate when followed by feedback about the DMTS response than when followed by no feedback (essentially the comparison between Conditions A and B in Experiment 1). Procedural differences may be responsible for this discrepancy, in particular the method used to control DMTS trial difficulty (time constraints in the previous research and stimulus number here), or the duration of conditions (our previous study used an extremely stringent stability criterion to guide condition changes, producing conditions lasting as much as 10 times longer than those reported here). Because feedback and consequence manipulations were not the main focus of this investigation, however, it is necessary to leave the resolution of these issues to future studies.

GENERAL DISCUSSION

The main purpose of this investigation was to determine whether the accuracy of humans' self-reports about DMTS performance would be influenced by the number of elements in the DMTS sample-stimulus compound. Generally speaking, it was. In this respect, the present results systematically replicate previous research in which self-report accuracy decreased as the difficulty of the DMTS target task was increased via time limits on DMTS responding (rather than via stimulus characteristics, as here). The first part of this general discussion reconsiders data generated in Experiment 1 to examine more closely some possible parallels across this study and the previous ones. The second part briefly considers, as a source of guidance for future studies, some research traditions with which the present studies seem to share emphases or procedures.

Possible Generalities Across Studies

The time-limit manipulation of previous studies and the sample-stimulus manipulation of the present investigation produced decreases in the percentage of DMTS responses that met a point contingency as well as decreases in the correspondence between self-reports and DMTS performance. Thus, it may be possible to treat DMTS success rate (percentage of successful DMTS responses) as a predictor of self-report accuracy, regardless of how the DMTS success rate is manipulated. Figure 6 examines this possibility. The top two rows of panels show the accuracy of self-reports as a function of DMTS success rates for each subject in Experiment 1. Each data point represents one trial type (based on the number of DMTS sample elements) from one experimental condition. In general, self-report accuracy was positively related to DMTS success.

The bottom row of panels in Figure 6 shows data from 3 subjects in our two previous self-report studies (Critchfield & Perone, 1990a, 1990b) in which DMTS difficulty was manipulated via time constraints (the figure omits a 4th subject who exhibited an unusual pattern of reporting that precludes meaningful comparison; see Critchfield & Perone, 1990a). Several other procedural details differed across studies, but all studies involved yes-no self-reports about DMTS success under a conjunctive speed-accuracy reinforcement contingency. Each data point in the bottom row of panels represents trials at one DMTS time limit in one experimental condition. The general pattern is the same as in the present Experiment 1: Self-reports were less accurate as DMTS success became less frequent.

Stimulus Control of Self-Reports By DMTS Response Characteristics

Figure 6 places results from three studies in a common context, but in the process treats DMTS success (point production) as a discrete event, although it actually depended on the conjunction of two response characteristics—speed and accuracy. The extent to which these two response characteristics were related to

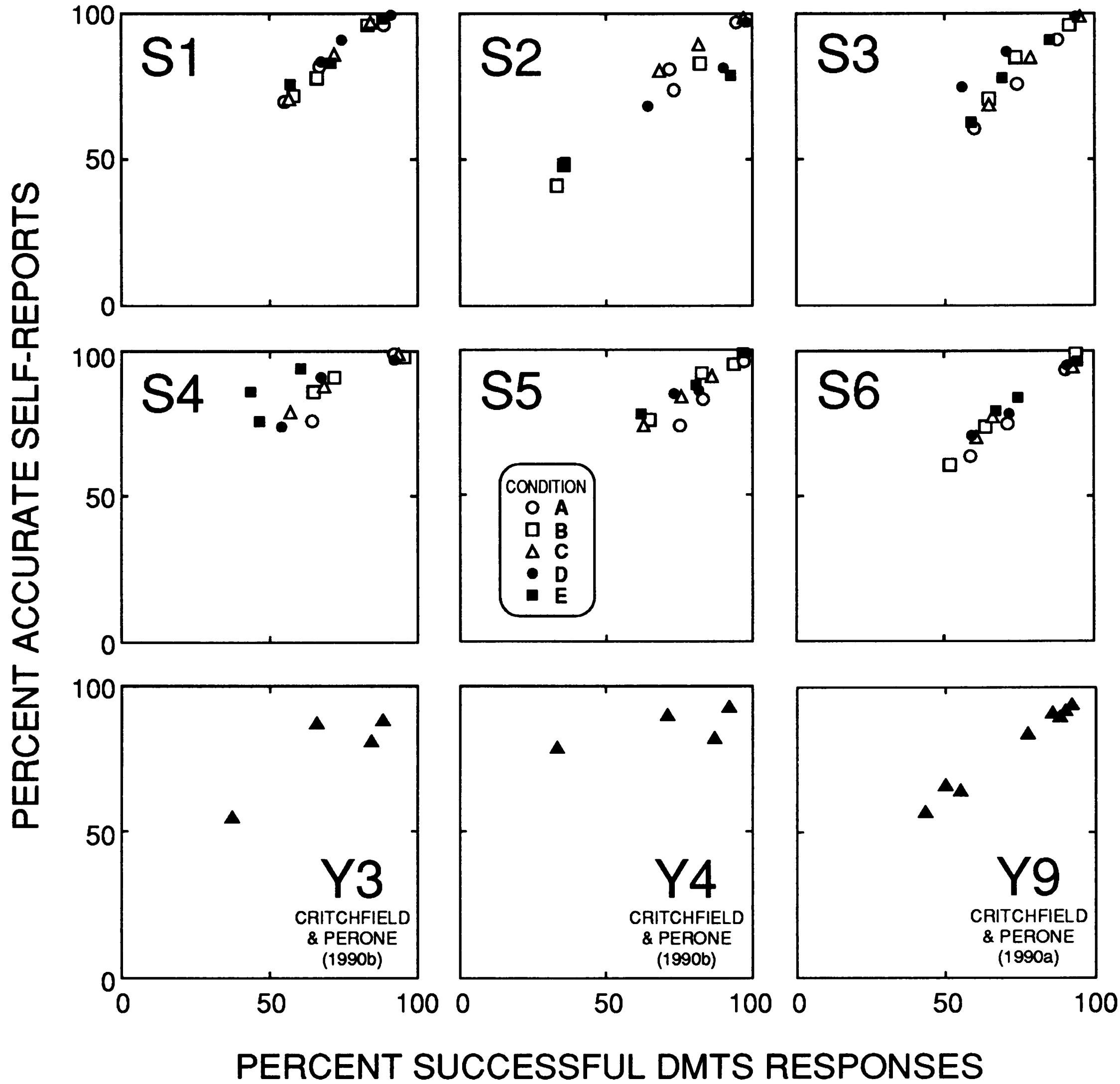


Fig. 6. Self-report accuracy as a function of DMTS success (percentage of responses that met the conjunctive speed-plus-accuracy reinforcement contingency). The top two rows of panels show data from Experiment 1, with experimental conditions keyed to the legend. The bottom row of panels shows data from two previous studies in which college students reported about DMTS success, and is adapted from Critchfield and Perone (1990b).

patterns of self-reporting may provide a second context in which results of different studies can be compared. An early experiment (Critchfield & Perone, 1990b) found that self-reports about DMTS success, similar to those investigated in the present research, were related to both the speed and accuracy of preceding DMTS responses but did not indicate whether these response characteristics exerted *differential* control over the reports. In a subsequent experiment (Critchfield & Perone, 1990a), when subjects were asked to report specifically about whether the preceding DMTS response was correct or faster than a time limit, they appeared to detect errors of accuracy more consistently than errors of speed, although this pattern was not precisely quantified.

Figures 7 and 8 present analyses of data from the present Experiment 1 that shed light on the relative stimulus control exerted by response speed and response accuracy over self-

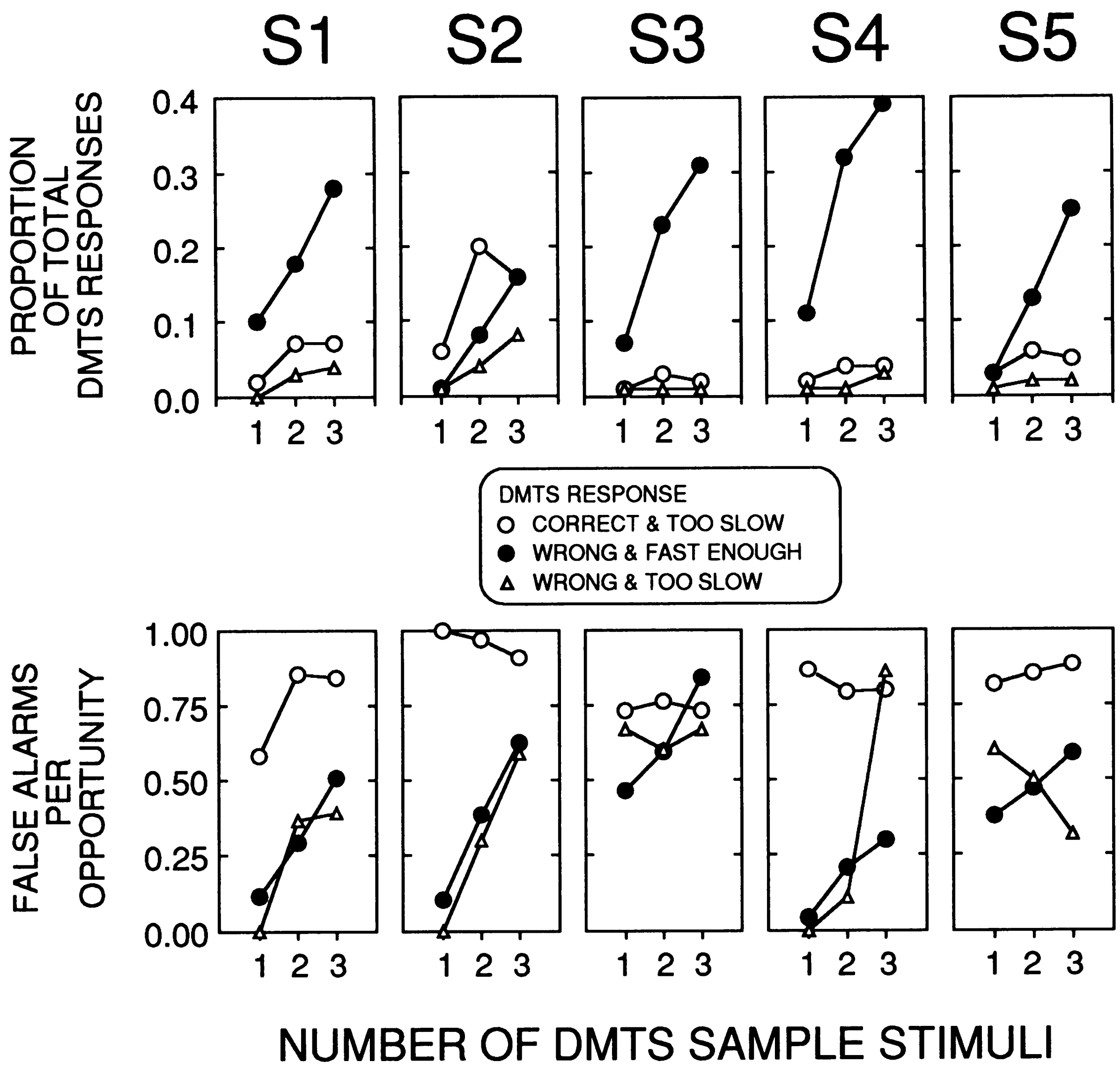


Fig. 7. Relative control of self-reports in Experiment 1 by the accuracy and speed of DMTS responses. Top: proportion of DMTS responses that were unsuccessful due to errors of accuracy, speed, or both, indicating the relative frequency of three situations in which false alarms could occur. Bottom: false alarms per opportunity following DMTS responses that were unsuccessful due to errors of accuracy, speed, or both. Each panel shows performance by 1 subject as a function of the number of elements in a compound DMTS stimulus.

reports of overall DMTS success. The analyses focus on rates of the more common type of inaccurate self-report, false alarms, in considering whether inaccurate self-reports about DMTS success were more closely correlated with DMTS speed or DMTS accuracy. (By definition, false alarms occurred after an error of DMTS response speed or response accuracy had been committed, and misses occurred only following successful DMTS responses; thus, only false alarm data can shed light on the relative stimulus control by DMTS response characteristics.)

Each panel in Figures 7 and 8 shows 1 subject's performance as a function of the number of sample stimuli present on each trial. Data have been summed across feedback and consequence conditions prior to analysis. Thus, the number of trials at each level of sample-stimulus complexity equals about 160 times

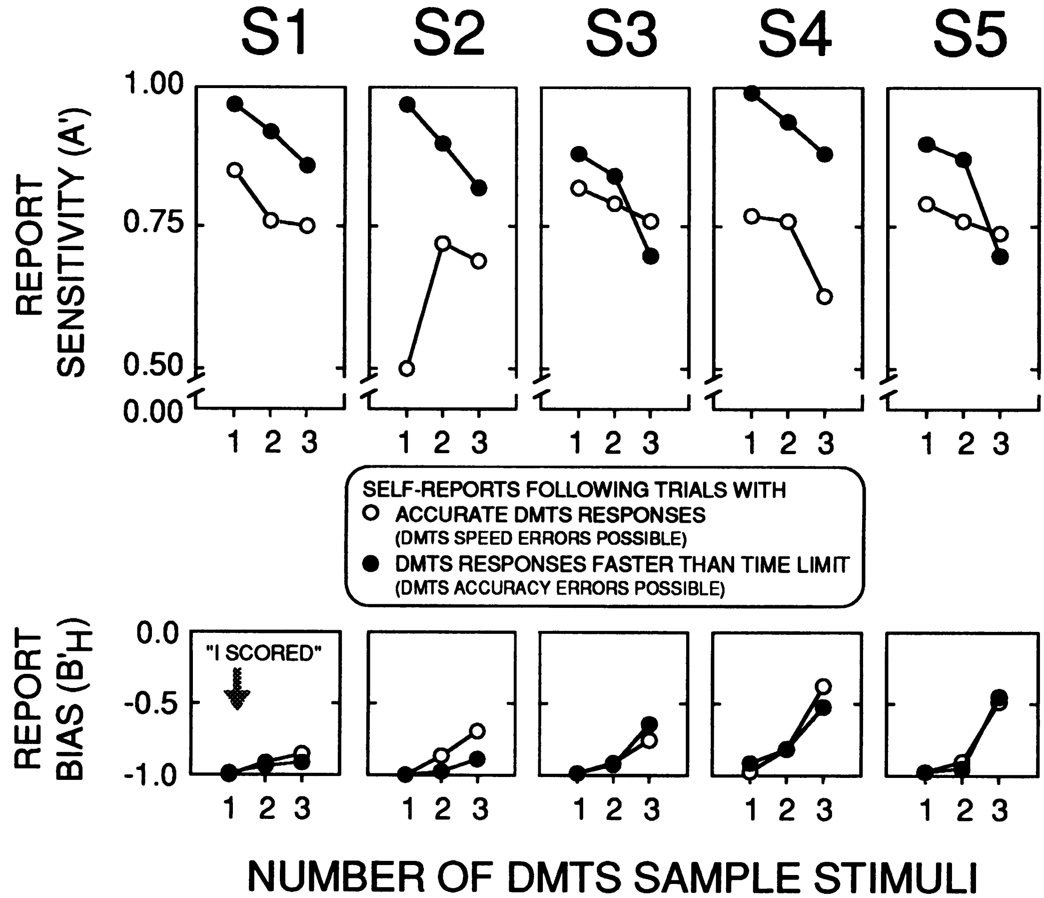


Fig. 8. Self-report sensitivity (top) and self-report bias (bottom) in Experiment 1, as influenced by DMTS accuracy errors and DMTS speed errors. See text for details of the calculation of sensitivity and bias scores.

the number of experimental conditions completed by each subject (a minimum of about 800 trials for each subject).

DMTS speed and accuracy errors. The top row of panels in Figure 7 shows how opportunities to make false alarms varied as a function of the sample-stimulus manipulation for 5 subjects (data for Subject 6 were not stored in a form amenable to this analysis). Separate functions show the percentage of DMTS responses that were incorrect, slower than the time limit, and both incorrect and too slow. For 4 of 5 subjects, unsuccessful DMTS responses were most often due to selection of an incorrect comparison stimulus; these errors of accuracy became more common as sample stimuli became more numerous. By contrast, responses that were unsuccessful due solely to a latency longer than the time limit typically were less common. Only rarely were DMTS responses both incorrect and too slow. Thus, although a speed-accuracy trade-off is expected with this type of target behavior (e.g., see Baron & Menich, 1985b), the sample-stimulus manipulation tended to have a greater impact on DMTS accuracy than on DMTS speed.

We have reported previously that manipulating the time limit on DMTS responding (2,000 vs. 500 ms) reduced DMTS accuracy more than DMTS speed (Critchfield & Perone, 1990a, 1990b). In the present study, a similar outcome was obtained when the time limit was held constant and the number of sample stimuli was manipulated (Figure 7, top). Thus, similar patterns of self-report accuracy may have occurred at least partly because structurally dissimilar manipulations created similar speed-accuracy trade-offs in the DMTS target performance. In other words, the top row of panels in Figure 7 provides reason to regard the time-limit and sample-stimulus manipulations as functionally similar.

False alarm rates controlled by DMTS speed versus accuracy. For purposes of estimating relative stimulus control over self-reports by DMTS response characteristics, the bottom row of panels in Figure 7 shows false alarms per opportunity as a function of DMTS response accuracy and DMTS response speed. The critical comparison is between failures to report errors of accuracy and failures to report errors of speed; failures to report DMTS responses that were both inaccurate and too slow provide only ambiguous information regarding relative stimulus control of self-reports. The figure shows that false alarms were especially likely when DMTS responses did not meet the time limit. For 4 of 5 subjects, false alarms were less likely when DMTS responses were inaccurate, although rates of this type of false alarm generally increased with the number of sample stimuli.

Figure 8 shows how patterns of false alarms depicted in Figure 7 affected the sensitivity and bias of self-reports. Grier's (1971) A′ and B'_H indices were recalculated after omitting either trials on which the DMTS response was inaccurate (meaning that only errors of speed could occur) or trials on which the DMTS response was too slow (meaning that only errors or accuracy could occur). Open circles show sensitivity or bias as affected by the speed of DMTS responses, and filled circles show sensitivity or bias as affected by the accuracy of DMTS responses. Each panel shows the relevant indices for 1 subject as a function of the number of DMTS sample stimuli.

In the top row of panels in Figure 8 it is evident that self-report sensitivity usually was higher when response accuracy was at issue than when response speed was at issue. This pattern is consistent with false alarm rates (per opportunity) shown in the bottom row of panels in Figure 7. The bottom row of panels in Figure 8 shows that self-report bias scores typically did not vary systematically as a function of DMTS speed versus accuracy. Taken together, Figure 7 (bottom) and Figure 8 (top) suggest that self-reports about overall DMTS success were better controlled by DMTS response accuracy than by DMTS response speed. This outcome is consistent with the finding in an earlier study that subjects reporting specifically about DMTS response speed and response accuracy detected their accuracy better than their speed (Critchfield & Perone, 1990a). The data from Experiment 1 thus provide an additional reason to draw parallels across studies.

Possible Discrepancies Across Studies

Whether apparent similarities across manipulations altering target-response success will hold up under further experimental scrutiny is not known, but an additional analysis provides at least one reason for skepticism. Figure 9 (top two rows of panels) shows self-report sensitivity and self-report bias scores from Experiment 1 as a function of DMTS success rates engendered by the sample-stimulus manipulation. Each panel is similar in format to those in Figure 6, except that there are two functions, one for sensitivity and one for bias. The data show substantial variability, but two general patterns are suggested. First, self-report sensitivity tended to increase as DMTS success became more frequent, producing the positively sloped functions in the upper right quadrant of each panel. Second, bias scores, though less orderly, generally approached the negative end of the scale (indicating a bias for reporting DMTS success), and became less extreme only where DMTS success was less frequent.

The bottom row of panels in Figure 9 shows data from the 3 subjects in our previous studies in which DMTS success rates were manipulated via the time limits on DMTS responding. Relations between DMTS success rates and self-report accuracy were similar across studies (Figure 6), but the same may not be true for relations between DMTS success rates and the sensitivity and bias of self-reports. For subjects in our previous studies, sensitivity did not change systematically with DMTS success rates, and bias scores covaried with success rates only for Subject Y4. It is tempting, but premature, to presume that the differences derive from manipulating time constraints instead of stimulus characteristics in the DMTS task. Multiple procedural differences across the studies make it impossible to pinpoint the source of this discrepancy without further experimentation. In the interim, however, the

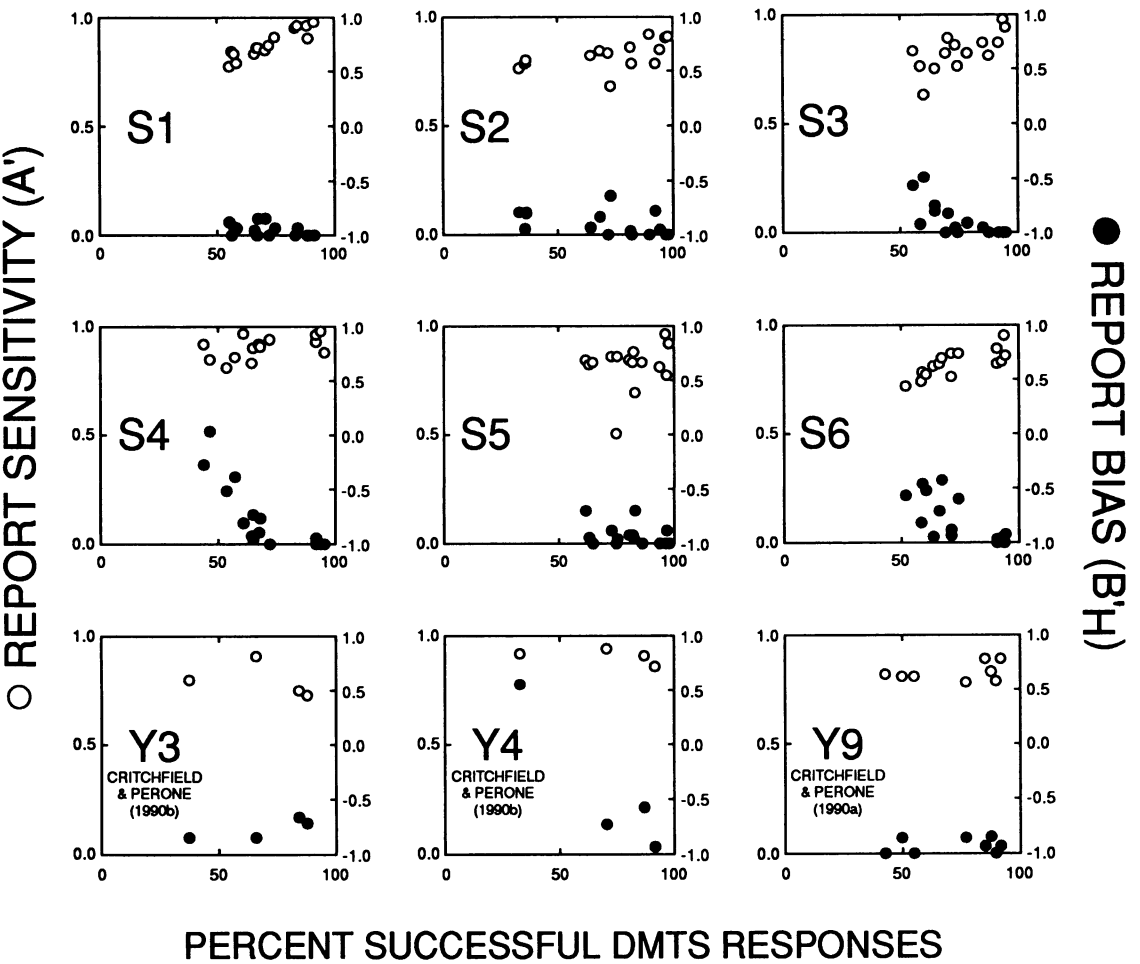


Fig. 9. Self-report sensitivity and self-report bias as a function of DMTS success (percentage of responses that met the conjunctive speed-plus-accuracy reinforcement contingency). The top two rows of panels show data from Experiment 1. The bottom row of panels shows data from two previous studies in which college students reported about DMTS success. Sensitivity scores (open circles) are scaled on the left ordinate. Bias scores (filled circles) are scaled on the right ordinate, with negative scores indicating a bias for reporting success and positive scores indicating a bias for reporting failure.

functions in Figure 9 are valuable in highlighting the limits of accuracy scores as dependent measures in studies of self-reporting: The patterns in Figure 9 could be neither derived nor predicted from those shown in Figure 6.

Directions for Future Research

Pervasive bias for reporting success. The pervasive bias for reporting DMTS success, alluded to in our previous self-report studies and quantified here (Figures 3, 8, and 9), bears further inspection. On a descriptive level, it remains to be seen whether this bias would hold across a broader range of parametric manipulations. For example, in Experiment 1, as DMTS success became less frequent bias scores became less extreme, but no subject showed a substantial bias for reporting failure. It is not known whether lower rates of DMTS success than those created here would be sufficient to produce reliable biases for reporting failure. In this regard, it may be significant that the sample-stimulus manipulation rarely produced DMTS success rates much lower than 50%.

At present, we cannot specify the origin of the bias for reporting successful responses. It may be idiosyncratic to, or an artifact of, our procedure, although reason exists to suspect otherwise. A substantial literature is concerned with the tendency of individuals completing psychological inventories and clinical assessments to overreport desirable characteristics and underreport undesirable ones (e.g., Arias & Beach, 1987; Best & Best, 1975; Borkenau & Ostendorf, 1989; Halbreich, Bakhai, Bacon, Goldstein, Asnis, Endicott, & Lesser, 1989; Hultsman, Hultsman, & Black, 1989). These "social desirability effects" may be consistent with self-reporting patterns described here; effects sometimes labeled as "self-serving" or "positive presentation" biases may be relevant as well. In some literatures, these patterns are believed to reflect a generic reporting bias rather than situational effects, and thus to manifest themselves in diverse circumstances (Brown, 1986; Furnham, 1986; McCormick, Walkey, & Green, 1986). Possibly, then, our procedures detected a bias that was extraexperimental rather than artifactual. If so, the bias should prove resistant to modifications of our present procedures (e.g., variations in preliminary training routine, different wordings of the self-report query, different types of target behavior, etc.).

Contact with cognitive literatures. The present investigation may be viewed as procedurally similar to research of current interest to cognitive psychologists studying human memory. Recent studies have focused on "metamemory," which typically is measured via self-reports about memory performance, and "activity memory," which typically is measured via self-reports about motor performance (e.g., Janowsky, Shimamura, & Squire, 1989; Kausler & Phillips, 1988). The methods employed in these studies rarely lend themselves to analyses like those presented here (but see Nelson, 1984). Nevertheless, to the extent that behavioral studies of self-reporting constitute a point of contact with other research traditions, considerable cross-pollination of ideas may be possible. Consider that in the present investigation the sample-stimulus manipulation would be said, in the parlance of cognitive memory research, to alter variables acting at the moment when stimulus information is *encoded* into memory. From this perspective, the question might be raised whether similar results would be obtained if the same type of manipulation took place at the moment when information is *retrieved* from memory, because many cognitive theories hold that encoding and retrieval are partially independent processes. Thus, a reasonable test of the generality of the present findings, manipulating the number of DMTS *comparison* stimuli instead of the number of sample elements, follows more readily from a cognitive perspective than a behavioral one (e.g., see Flavell & Wellman, 1977). It seems safe to predict that increasing the number of comparison stimuli would decrease DMTS success, but it is not known how self-reports about that success would be affected. Fortunately, unlike many procedures currently popular in cognitive studies of human metamemory, the procedures used here allow the comparison to be made because precisely the same manipulation is possible at both "encoding" and "retrieval."

Contact with psychophysical literatures. The procedures employed here to study self-reports also bear structural similarity to those used in psychophysical research to study verbal reports of external stimuli. Signal-detection analyses have contributed to psychophysical research in part by clarifying the extent to which reports of external stimuli are behavior under environmental control (Gescheider, 1985; Goldiamond, 1964). Research addressing the signal-detection properties of verbal self-reports might therefore look to psychophysical studies for guidance, especially in terms of how various types of consequences, current or historical, could be expected to influence reporting biases (e.g., Davison & Tustin, 1978). The psychophysical literature might be especially useful in suggesting studies to explain why a point contingency could reduce self-report accuracy in Experiment 2 but fail to improve it in Experiment 1. Initially, however, parallels between self-reports and reports of external stimuli should be drawn with some caution. Typically in psychophysical research the physical properties of stimuli that provide the basis for detection by observers (e.g., light intensity or wavelength, vibration amplitude, stimulus probability, etc.) are well understood and therefore manipulable in minute increments. By contrast, a primary methodological challenge in a "psychophysics of self-observation" would be the identification and control of those aspects of target behavior that constitute the antecedent stimulus for self-reports. In the present research, for example, self-reports by different "self-observers," although similar in overall correspondence to target-response success, were controlled to different degrees by two broad aspects of that success, speed and accuracy (each of which might also turn out to be multidimensional). Until relevant target-response dimensions are better understood, it could prove difficult to control target behavior for self-reports with the same precision that psychophysical researchers control the occurrence and characteristics of external stimuli such as lights and tones.

Conclusions

As noted in the introduction, the answers to broad questions surrounding the veracity of self-reports are beyond the scope of any single investigation and, indeed, probably of any single line of research. Behavioral approaches to the study of self-reports are relatively new, given the long history of self-reports in experimental psychology (e.g., see Boring, 1953), but encouragement may be drawn from situations in which the data are orderly and relevant to findings in other research traditions. The present studies appear to qualify as useful in this regard.

REFERENCES

Appel, J. B., & Dykstra, L. A. (1977). Drugs, discrimination, and signal detection theory. In T. Thompson & P. B. Dews (Eds.), *Advances in behavioral pharmacology* (Vol. 1, pp. 139–166). New York: Academic Press.

Arias, I., & Beach, S. R. (1987). Validity of self-reports of marital violence. *Journal of Family Violence,* **2,** 139–149.

Baron, A., & Menich, S. R. (1985a). Age-related effects of temporal contingencies on response speed and memory: An operant analysis. *Journal of Gerontology,* **40,** 60–70.

Baron, A., & Menich, S. R. (1985b). Reaction times of older and younger men: Effects of compound samples and a prechoice signal on delayed matching-to-sample performances. *Journal of the Experimental Analysis of Behavior,* **44,** 1–14.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Reviews,* **74,** 183–200.

Bernstein, D. J., & Michael, R. L. (1990). The utility of verbal and behavioral assessments of value. *Journal of the Experimental Analysis of Behavior,* **54,** 173–184.

Best, J. A., & Best, H. (1975). Client self-monitoring in clinical decision-making. *Canada's Mental Health,* **23,** 9–11.

Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality,* **3,** 31–45.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition,* **4,** 353–376.

Caracci, G., Mukherjee, S., Roth, S. D., & Decina, P. (1990). Subjective awareness of abnormal involuntary movements in chronic schizophrenic patients. *American Journal of Psychiatry,* **147,** 295–298.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior.* New York: Springer-Verlag.

Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior,* **38,** 233–248.

Ciminero, A. R., Nelson, R. O., & Lipinski, D. P. (1977). Self-monitoring procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 195–232). New York: Wiley.

Critchfield, T. S., & Perone, M. (1990a). Verbal self-reports as a function of speed, accuracy, and reinforcement of the reported performance. *Psychological Record,* **40,** 541–554.

Critchfield, T. S., & Perone, M. (1990b). Verbal self-reports of delayed matching to sample by humans. *Journal of the Experimental Analysis of Behavior,* **53,** 321–344.

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal detection theory. *Journal of the Experimental Analysis of Behavior,* **29,** 331–336.

de Freitas Ribeiro, A. (1989). Correspondence in children's self-report: Tacting and manding aspects. *Journal of the Experimental Analysis of Behavior,* **51,** 361–367.

Elaand, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology,* **26,** 442–451.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail, Jr. & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–33). Hillsdale, NJ: Erlbaum.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences,* **7,** 385–400.

Gardner, R. M., Martinez, R., & Espinoza, T. (1987). Psychophysical measurement of body image of self and others in obese subjects. *Journal of Social Behavior and Personality,* **2,** 205–217.

Gescheider, G. A. (1985). *Psychophysics: Method, theory, and application.* Hillsdale, NJ: Erlbaum.

Goldiamond, I. (1964). Response bias in perceptual communication. In D. M. Rioch & E. A. Weinstein (Eds.), *Disorders of communication* (pp. 334–363). Baltimore: Williams and Wilkins.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin,* **75,** 424–429.

Halbreich, U., Bakhai, Y., Bacon, K. B., Goldstein, S., Asnis, G. M., Endicott, J., & Lesser, J. (1989). The normalcy of self-proclaimed "normal volunteers." *American Journal of Psychiatry,* **146,** 1052–1055.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General,* **108,** 356-388.

Hayes, S. C. (1986). The case of the silent dog—Verbal reports and the analysis of rules: A review of Ericsson and Simon's *Protocol Analysis: Verbal Reports As Data. Journal of the Experimental Analysis of Behavior,* **45,** 351-363.

Hefferline, R. F., & Perera, T. B. (1963). Proprioceptive discrimination of a covert operant without its observation by the subject. *Science,* **139,** 834-835.

Hosseini, J., & Ferrell, W. R. (1982). Detectability of correctness: A measure of knowing that one knows. *Instructional Science,* **11,** 113-127.

Hultsman, W. Z., Hultsman, J. T., & Black, D. R. (1989). "Response peaks" as a component of measurement error: Assessment implications for self-reported data in leisure research. *Journal of Leisure Research,* **21,** 310-315.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology,* **17,** 3-11.

Kagan, J. (1988). The meanings of personality predicates. *American Psychologist,* **43,** 614-620.

Kausler, D. H., & Phillips, P. L. (1988). Instructional variation and adult age differences in activity memory. *Experimental Aging Research,* **14,** 195-199.

McCormick, I. A., Walkey, F. H., & Green, D. E. (1986). Comparative perceptions of driver ability—A confirmation and expansion. *Accident Analysis and Prevention,* **18,** 205-208.

Natsoulis, T. (1988). Is any state of consciousness self-intimating? *Journal of Mind and Behavior,* **9,** 167-203.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin,* **95,** 109-133.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* **84,** 231-259.

Overton, D. A. (1987). Applications and limitations of the drug discrimination method for the study of drug abuse. In M. A. Bozarth (Ed.), *Methods of assessing the reinforcing properties of abused drugs* (pp. 291-340). New York: Springer-Verlag.

Risley, T. R., & Hart, B. (1968). Developing correspondence between the non-verbal and verbal behavior of preschool children. *Journal of Applied Behavior Analysis,* **1,** 267-281.

Shimoff, E. (1986). Post-session verbal reports and the experimental analysis of behavior. *Analysis of Verbal Behavior,* **4,** 19-22.

Shimp, C. P. (1983). The local organization of behavior: Dissociations between a pigeon's behavior and self-reports of that behavior. *Journal of the Experimental Analysis of Behavior,* **39,** 61-68.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review,* **52,** 291-294.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Tucker, J. A., Vuchinich, R. E., Harris, C. V., Gavornik, M. G., & Rudd, E. J. (1991). Agreement between subject and collateral verbal reports of alcohol consumption in older adults. *Journal of Studies on Alcohol,* **52,** 148-155.

Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior,* **46,** 15-35.

Ziriax, J. M., & Silberberg, A. (1978). Discrimination and emission of different key-peck durations in the pigeon. *Journal of Experimental Psychology: Animal Behavior Processes,* **4,** 1-21.

Zuroff, D. C., Colussy, S. A., & Wielgus, M. S. (1983). Selective memory and depression: A cautionary note concerning response bias. *Cognitive Therapy and Research,* **7,** 223-231.

Received January 30, 1992
Final acceptance August 25, 1992